

Method for Transmitting Audio Signals According to the  
Prioritizing Pixel Transmission

The invention relates to a method for transmitting audio signals according to the prioritizing pixel transmission according to the preamble of patent claim 1.

Currently a multiplicity of methods exists for the compressed transmission of audio signals. Essentially the following methods are among them:

- Reduction of the sampling rate, for example 3 kHz instead of 44 kHz
- Nonlinear transmission of the sampled values, for example in ISDN transmission
- Utilization of previously stored acoustic sequences, for example MIDI or voice simulation
- Employing Markov models for the correction of transmission errors.

The commonalities of the known methods reside therein that even at lower transmission rates satisfactory voice intelligibility is still provided. This is substantially attained through the formation of mean values. However, different voices of the source yield similarly sounding voices in the [rate] lowering, such that, for example voice fluctuations, which are detectable in normal conversation, are no longer transmitted. This results in a marked restriction in the quality of communication.

Methods for compressing and decompressing of image or video data by means of prioritized pixel transmission are described in the applications DE 101 13 880.6 (corresponding to PCT/DE02/00987) and DE 101 52 612.1 (corresponding to PCT/DE02/00995). In these methods, for example digital image or video data are processed, which are comprised of an array of individual pixels, each pixel comprising a pixel value varying in time, which describes color or brightness information of the pixel. According to the invention, to each pixel or each pixel group a priority is assigned and the pixels are stored corresponding to their prioritization in a priority array. This array contains at each point in time the pixel values sorted according to

prioritization. These pixels and the pixel values utilized for the calculation of the prioritization are transmitted or stored corresponding to the prioritization. A pixel receives a high priority if the differences to its adjacent pixels are very large. For the reconstruction the particular current pixel values are represented on the display. The pixels not yet transmitted are calculated from the already transmitted pixels. These methods can in principle also be utilized for the transmission of audio signals.

The invention therefore has at its aim to specify a method for transmitting audio signals, which operates with minimum losses even at low transmission bandwidths.

This aim is attained according to the invention through the characteristics of patent claim 1.

According to the invention the audio signal is first resolved into a number  $n$  of spectral components. The resolved audio signal is stored in a two-dimensional array with a multiplicity of fields, with frequency and time as the dimensions and the amplitude as the particular value to be entered in the field. Subsequently from each individual field and at least two fields adjacent to this field of the array, groups are formed, and to the individual groups a priority is assigned, the priority of a group being selected higher the greater the amplitudes of the group values are and/or the greater the amplitude differences of the values of a group are and/or the closer the group is to the current time. Lastly, the groups are transmitted to the receiver in the sequence of their priority.

The new method essentially rests on the foundations of Shannon. According to them, the signals can be transmitted free of loss if they are sampled at the twofold frequency. This means that the sound can be resolved into individual sinusoidal oscillations of different amplitude and frequency. Accordingly, the acoustic signals can be unambiguously restored without losses by transmitting the individual frequency components, including amplitudes and phases. Herein is in particular utilized that the frequently occurring sound sources, for example musical instruments or the human voice, are comprised of resonance bodies, whose resonant frequency does not change at all or only slowly.

Advantageous embodiments and further developments of the invention are specified in the dependent patent claims.

An embodiment example of the invention will be described in the following. Reference shall be made in particular also to the specification and the drawing of the earlier patent applications DE 101 13 880.6 and DE 101 52 612.1.

First, the sound is picked up, converted into electric signals and resolved into its frequency components. This can be carried out either through FFT (Fast Fourier Transformation) or through  $n$ -discrete frequency-selective filters. If  $n$ -discrete filters are utilized, each filter picks up only a single frequency or a narrow frequency band (similar to the cilia in the human ear). Consequently, there is at each point in time the frequency and the amplitude value at this frequency. The number  $n$  can assume different values according to the end device properties. The greater  $n$  is, the better the audio signal can be reproduced.  $n$  is consequently a parameter with which the quality of the audio transmission can be scaled.

The amplitude values are placed into intermediate storage in the fields of a two-dimensional array.

The first dimension of the array corresponds to the time axis and the second dimension to the frequency. Therewith every sampled value with the particular amplitude value and phase is unambiguously determined and can be stored in the associated field of the array as an imaginary number. The voice signal is consequently represented in three acoustic dimensions (parameters) in the array: the time for example in milliseconds (ms), perceptually discerned as duration as the first dimension of the array, the frequency in Hertz (Hz), perceptually discerned as tone pitch, as the second dimension of the array and the energy (or intensity) of the signal, perceptually discerned as volume or intensity, which is stored as a numerical value in the corresponding field of the array.

In comparison to the applications DE 101 13 880.6 and DE 101 52 612.1, the frequency corresponds for example to the image height, the time to the image width and the amplitude of the audio signal (intensity) to the color value.

Similar to the method of the prioritizing of pixel groups in image/video coding, groups are formed of adjacent values and these are prioritized. Each field, considered by itself, together with at least one, preferably however several adjacent fields form one group. The groups are comprised of the position value, defined by time and frequency, the amplitude value at the position value, and the amplitude values of the allocated values corresponding to a previously defined form (see Figure 2 of applications DE 101 13 880.6 and DE 101 52 612.1). Especially those groups receive a very high priority which are close to the current time and/or whose amplitude values, in comparison to the other groups, are very large and/or in which the amplitude values within the group differ strongly. The pixel group values are sorted in descending order and stored or transmitted in this sequence.

The width of the array (time axis) preferably has only a limited extent (for example 5 seconds), i.e. only signal sections of, for example, 5 seconds length are always processed. After this time (for example 5 seconds) the array is filled with the values of the succeeding signal sections.

The values of the individual groups are received in the receiver according to the above described prioritization parameters (amplitude, closeness of position in time and amplitude differences from adjacent values).

In the receiver the groups are again entered into a corresponding array. According to patent applications DE 101 13 880.6 and DE 101 52 612.1, subsequently from the transmitted groups the three-dimensional spectral representation can again be generated. The more groups were received, the more precise is the reconstruction. The not yet transmitted array values are calculated by means of interpolation from the already transmitted array values. From the thus generated array, subsequently in the receiver a corresponding audio signal is generated which subsequently can be converted into sound.

For the synthesis of the audio signal for example  $n$  frequency generators can be utilized, whose signals are added to an output signal. Through this parallel structuring of  $n$  generators good scalability is attained. In addition, the clock rate can be drastically

reduced through parallel processing, such that, due to a lower energy consumption, the playback time in mobile end devices is increased. For parallel application for example FPGAs or ASICs of simple design can be employed.

The described method is not limited to audio signals. The method can be effectively applied in particular where several sensors (sound sensors, light sensors, tactile sensors, etc.) are utilized, which continuously measure signals which subsequently can be represented in an array (of  $n$ th order).

The advantages compared to previous systems reside in the flexible applicability in the case of increased compression rates. By utilizing an array which is supplied from different sources, the synchronization of the sources is automatically obtained. The corresponding synchronization in conventional methods must be ensured through special protocols, or measures. In particular in video transmission with long propagation times, for example satellite connections, where sound and image are transmitted across different channels, frequently a lacking synchronization of the lips with the voice is noticeable. This can be eliminated through the described method.

Since the same fundamental principle of the prioritizing pixel group transmission can be utilized in voice, image and video transmission, a strong synergy effect is utilizable in the implementation. In addition, in this way the simple synchronization between language and images can take place. In addition, there could be arbitrary scaling between image and audio resolution.

If an individual audio transmission according to the new method is considered, in terms of voice a more natural reproduction results, since the frequency components (groups) typical for each human being are transmitted with highest priority and therewith free of loss.